

University of Dundee

## Adaptive Subtitles

Gorman, Benjamin M.; Crabb, Michael; Armstrong, Mike

*Published in:*

CHI 2021 - Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems

*DOI:*

[10.1145/3411764.3445509](https://doi.org/10.1145/3411764.3445509)

*Publication date:*

2021

*Document Version*

Peer reviewed version

[Link to publication in Discovery Research Portal](#)

*Citation for published version (APA):*

Gorman, B. M., Crabb, M., & Armstrong, M. (2021). Adaptive Subtitles: Preferences and Trade-Offs in Real-Time Media Adaption. In *CHI 2021 - Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems: Making Waves, Combining Strengths* (pp. 1-11). [733] (Conference on Human Factors in Computing Systems - Proceedings). Association for Computing Machinery. <https://doi.org/10.1145/3411764.3445509>

### General rights

Copyright and moral rights for the publications made accessible in Discovery Research Portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from Discovery Research Portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain.
- You may freely distribute the URL identifying the publication in the public portal.

### Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

# Adaptive Subtitles: Preferences and Trade-Offs in Real-Time Media Adaption

Benjamin M. Gorman  
Bournemouth University  
Bournemouth, England, UK  
bgoman@bournemouth.ac.uk

Michael Crabb  
University of Dundee  
Dundee, Scotland, UK  
mzcrabb@dundee.ac.uk

Mike Armstrong  
BBC Reserach and Development  
Salford, England, UK  
mike.armstrong@bbc.co.uk

## ABSTRACT

Subtitles can help improve the understanding of media content. People enable subtitles based on individual characteristics (e.g., language or hearing ability), viewing environment, or media context (e.g., drama, quiz show). However, some people find that subtitles can be distracting and that they negatively impact their viewing experience. We explore the challenges and opportunities surrounding interaction with real-time personalisation of subtitled content. To understand how people currently interact with subtitles, we first conducted an online questionnaire with 102 participants. We used our findings to elicit requirements for a new approach called Adaptive Subtitles that allows the viewer to alter which speakers have subtitles displayed in real-time. We evaluated our approach with 19 participants to understand the interaction trade-offs and challenges within real-time adaptations of subtitled media. Our evaluation findings suggest that granular controls and structured onboarding allow viewers to make informed trade-offs when adapting media content, leading to improved viewing experiences.

## CCS CONCEPTS

• **Human-centered computing** → *Interaction paradigms*.

## KEYWORDS

Subtitles, Captions, Closed-captions, Media, Adaptive-Interfaces

### ACM Reference Format:

Benjamin M. Gorman, Michael Crabb, and Mike Armstrong. 2021. Adaptive Subtitles: Preferences and Trade-Offs in Real-Time Media Adaption. In *CHI Conference on Human Factors in Computing Systems (CHI '21)*, May 8–13, 2021, Yokohama, Japan. ACM, New York, NY, USA, 11 pages. <https://doi.org/10.1145/3411764.3445509>

## 1 INTRODUCTION

Subtitles (or closed-captions<sup>1</sup>) are used by viewers to help them understand and enjoy media content. A British Broadcasting Corporation (BBC) audience survey reported that 10% of television

viewers in the UK use subtitles daily, and 6% use subtitles "most of the time" [4]. It has also been reported that 18% of all BBC iPlayer content (i.e. online media streaming) is viewed with subtitles on, with this increasing to over 20% for tablet users [1]. With such a large percentage of media being consumed with subtitles to assist in the viewing experience, it is important to understand the reasons behind this usage. Developing an understanding into why viewers enable subtitles may allow the overall viewing experience to be better tailored for audience members on an individual basis.

It has been suggested that 80% of subtitle users do not have a hearing impairment [49]. Despite this, research involving subtitles commonly focuses on its usage as an accessibility feature. In this work, we focus on understanding how viewers adapt their subtitle usage depending on the content and context they are watching under. Specifically, we aim to understand the challenges and opportunities within personalising subtitle interactions.

To achieve our aim, we first conducted an online questionnaire with 102 participants to explore subtitle interaction patterns and experiences. Our participants highlighted specific viewing challenges surrounding the language being spoken, accents, scene-context, and programme quality. Participants also described turning on subtitles for specific accents, actors, or content types, and that they had to interrupt their viewing to do so.

Previous work has focused on adaptive interfaces as a system-wide adoption. Based on our survey findings, we determined that the content being consumed is an additional factor that should be considered within such interfaces. With this in mind, we introduce a new approach called Adaptive Subtitling that allows subtitles to be adapted based on the viewers' individual preferences. To evaluate our approach, we created a system that gives the viewer control over which individual speakers have subtitles enabled. This allowed us to explore the opportunities and trade-offs that exist when allowing for real-time personalisation of media content.

**Paper Contributions:** This paper makes three contributions: First, we contribute online questionnaire data from 102 participants that provides an understanding into why people use subtitles, questioning the breadth of use cases that should be considered in their design. Second, we introduce Adaptive Subtitles, which allows real time adaptation of subtitled content by the user, and make available sample code for how this can be implemented using modern web technologies through a second screen application. Third, we evaluated Adaptive Subtitles through a lab based user study with 19 participants and introduce design considerations that outline the trade-offs involved when developing real-time media adaptations. For transparency, we provide anonymised participant data and project code as supplementary material attached to this work.

<sup>1</sup>Closed captions (CC) also provide a text description of sound effects. Most streaming sites only have the option for 'English[CC]' for English subtitles and therefore in this work we collectively refer to both as *subtitles*.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

CHI '21, May 8–13, 2021, Yokohama, Japan

© 2021 Copyright held by the owner/author(s). Publication rights licensed to ACM.  
ACM ISBN 978-1-4503-8096-6/21/05...\$15.00  
<https://doi.org/10.1145/3411764.3445509>

## 2 RELATED WORK

Subtitles are used to convey spoken dialogue and sound effects to the viewer within media content. Subtitles enable audiences to gain additional information about particular aspects of a film or television show that could relate to character identification, time markers, narrative summary, dialogue, and story commentary [14]. The most prominent use for subtitles is as an access service that enables people with hearing impairments to better understand media content. One of the main reasons in creating modern subtitled content is to assist viewers with hearing loss [16] and to prevent this group being “*shut out*” from broadcast media [38].

Despite *subtitles as an access service* [66] being one of the primary reasons for television content being subtitled, it is estimated that only 20% of people use subtitles for this reason [49]. There are many factors that determine whether an individual may choose to watch media content with subtitles turned on. Situational factors can influence why individuals may be unable to use traditional audio as the main method of understanding media content [17], and the reasons for using subtitles can be as unique as the individuals themselves that are using them [19]. Alternative reasons for subtitle usage may include characters mumbling, background noise in TV, watching in loud environment, having to have the sound low, and the use of unfamiliar words or accents [56]. Context, therefore, is key in understanding why an individual may, or may not, watch video content with subtitles enabled.

### 2.1 Impact of Subtitles on Viewing Experience

Subtitles are designed to have a positive impact within media, however sometimes their inclusion can lead to a reduction in overall viewing experience. It has been suggested that when subtitles are present they can take up ~37% of a users visual attention [12], and eye-tracking work has found that participants spent ~84% of their viewing time on the subtitles when watching media content [31]. It has been argued that the presence of subtitles within a movie can disengage the viewer from the experience of the film and have a negative impact on the overall enjoyment [60]. The presence of subtitles within 3D stereoscopic movies can negatively effect the visual comfort of a viewer [36], and the inclusion of subtitles to musical pieces can lower the amount of expression that a user perceives from a performance, with possible justification for this being the multi-tasking elements that are required to listen to music and read subtitles at the same time [57]. Subtitles, therefore, have the potential to distract viewers if they are present when not required.

One method that can be used to alter the impact that subtitles have in overall viewing experience is to adapt their position within the media content. The traditional position for subtitles is at the bottom of the media being presented, however subtitle placement can be changed to avoid obscuring content and to reduce distraction [16]. The display of subtitle text can also be adapted based on device size [29], and available space outside of the media content frame [19]. Subtitles can also be dynamically positioned [9], with this method showing potential in increasing the overall viewing experience of subtitled content [21] and also being important when considering placement in VR environments [28].

Despite efforts to create new methods of presenting subtitles, they can be distracting to viewers, with dwell time highest for those

not using them as an access service [12]. Viewers must perform a complex number of steps for each new subtitle block that appears, and use a variety of communication channels concurrently whilst doing so [34]. However, despite the reduction in viewing experience, the inclusion of subtitles has positive benefits outside of their usage as an accessibility aid [16]. The presence of same language subtitles may decrease cognitive load when used in an education setting [33] and it has been suggested they focus attention [33], which may be more important for lean-forward experiences [32].

### 2.2 Customising and Personalising Experiences

Developing experiences that match individuals’ preferences is a complex task that involves understanding user needs [58]. The overall experience of using a system is not based on the system itself, but more on the individual that is using it at a given point in time [59]. Creating services that cater for specific user needs is not a domain where one-size-fits-all due to the ever changing abilities that individuals may have when using a piece of technology and the environments that they may use these technologies in [26]. Personalisation as a method to increase the overall experience of a service is one that as shown promise in a number of media contexts.

Systems with adaptive accessibility require differing levels of user involvement. System-led adaptations typically involve user models to be created that facilitate adaptations automatically for a user [44]. User-led adaptations involve the users themselves leading the adaptations that are taking place in a proactive manner [22]. Both of these methods are valid and their usage depends on the user, context of use, and complexity of the interface and interactions being adapted.

Acceptance of customised subtitled content is not based on comprehension or readability but on culture, habits, age, attitudes, and content [39]; more commonly known as factors relating to User Experience (UX) [47]. Comfort, rather than readability, has been suggested as a metric to use when creating guidelines for subtitle positioning [65] and in this regard, participants respond positively when given the ability to personalise the position of subtitles when viewing online media [19]. The most common method of subtitle adaptation is based on language, with different subtitle tracks available to suit viewer preferences. The use of second language subtitling can be used within an education setting to improve word recognition [40] but can also cause confusion between dialects [43]. Many online streaming services (e.g. Netflix [45]) also allow the user to customise how subtitles visually appear across content.

Personalising media content is a difficult task due to the number of interlinking steps involved within the media creation process. However, recent advancements in the use of Object Based Media (OBM) has changed the way that media production and consumption can be thought about [3]. OBM retains content as component parts, rather than rendering a finished artefact, and delivers these separately to the viewer. This allows media to be presented to the viewer in a personalised manner that takes into account individual needs whilst keeping overall viewing experience as a key concept in content delivery [2, 20].

### 2.3 Understanding Personalised Subtitles

The way that audiences consume media content has shifted towards interactive web-based players [48]. As such, viewers now expect a

personalised viewing experience across all aspects of media, including subtitles. Technology usage is not a one-size-fits all domain [26] but many services do not embrace personalisation opportunities. Previous work has focused on adaptive interfaces as a system-wide adoption [67], we take this concept and hypothesise that the content that is being consumed is an additional factor that should be considered. We question why the experience of watching subtitled content is constrained to a binary choice when the content [7], context [51], and abilities of viewers differs significantly.

Typically, research has focused on adapting the appearance of subtitles with regards to location [16], position [9], and text size [29]. However, even if you change how subtitles appear, if they are present all of the time, and especially when viewers don't need them, they can be distracting [12, 31, 60].

Taking the above into consideration, we formulated this paper's main research question as **RQ1**: "What challenges and opportunities exist when interacting with real-time personalisation of subtitled media content". To answer this, we have the following aims:

**Aim 1:** To understand how people currently interact with subtitles. We achieve this by carrying out an online questionnaire to determine how people currently interact with subtitled media content.  
**Aim 2:** To understand trade-offs and challenges that exist when allowing for real-time personalisation of subtitled media content. We achieve this by the design and evaluation of an interaction method that allows for real-time personalisation of subtitled content.

### 3 STUDY 1: QUESTIONNAIRE ON SUBTITLE USAGE PATTERNS

To explore the context around when and why people choose to use subtitles, we conducted an online questionnaire with people who self-reported regularly watching scripted media. There were four questions framing our questionnaire: 1) How often do people watch scripted entertainment (e.g., movies, documentaries)? 2) What services do people use to watch scripted entertainment? 3) What type of content do people turn subtitles on for? 4) Why do people use subtitles when watching specific types of content?

#### 3.1 Design & Method

There were 24 questions across four sections. The first section contained nine questions that were used to gather basic demographic information; age, gender, highest level of education, level of computer literacy, and details surrounding the participants' hearing ability. The second section contained five questions and focused on participants' viewing frequency: "What devices do you watch scripted entertainment on?", "How often do you watch scripted entertainment?", "How many hours per day do you watch scripted entertainment?", "What services do you use to watch scripted entertainment?", "Alongside terrestrial TV and online streaming, do you use any of the following to watch scripted entertainment?".

The third section contained 10 questions and focused on participants' subtitle usage: "Do you regularly watch scripted entertainment with subtitles turned on?", "What are the reasons you watch scripted entertainment with subtitles turned on?", "How often do you watch scripted entertainment with subtitles turned ON?", "How often do you watch scripted entertainment with subtitles turned OFF?", "Have you ever turned subtitles on for a specific show or type

of content?", "If yes - What type of content do you turn subtitles on for?", "Have you ever needed to pause or stop watching a show because you couldn't hear what was being said on screen?", "If Yes, please explain:", "Have you ever had trouble understanding an accent on a programme?", "If Yes, please explain:".

Ethical approval for the questionnaire was obtained from our ERB. We distributed the questionnaire using social media (e.g., Facebook, Twitter), Reddit (r/samplesize), university mailing lists, and specific charities and organisations (e.g., [RNID](#)).

#### 3.2 Participants

In total, 102 participants completed the questionnaire. Participants were aged between 18 and 86 ( $M = 29.57$ ,  $SD = 13.18$ ), with one not given. We used an open text field for gender: Male = 50, Female = 47, Transgender Guy = 1, Other = 1<sup>2</sup>, Not Given = 3. Participants reported on their highest level of education: University (72 participants), College (9), High School (16), Other (4), and Not Given (1). Participants reported on their level of computer literacy: Excellent (80 participants), Good (19), Fair (3), and Poor (0).

In total, 24 participants self-reported having a hearing loss. Participants were asked to describe their hearing loss using textual descriptions provided by RNID [55]: Mild (14 participants), Moderate (5), Severe (4), and Profound (1). Participants also reported how long they had a hearing loss. This was an open text field, that was then categorised into '0-5 Years' (7 participants), '5-10 Years' (7), '10-15 Years' (0), '15-20 Years' (1) and '20 Years plus' (9). Participants were also asked to report the cause of the hearing loss. This was presented as checkboxes with an 'Other' field: Ageing (3 participants), Congenital (5), Viral Infection (4), Exposure to loud noise (6), Unknown (3), Ear Damage (1), Head Trauma (1), Otosclerosis (1), and Not Given (2). Participants reported if they used any assistive technology, with nine participants reporting using hearing aids.

### 4 QUESTIONNAIRE FINDINGS

Closed-ended questions are reported by frequency of responses. Open-ended questions were analysed independently using open coding [63], based on existing procedure [61, 62]. We used the following four-step process:

- (1) **Generating and collating initial codes:** The lead author read all responses, taking note of initial codes. These were generated using a data-driven approach, collated, collapsed and developed into an initial codebook.
- (2) **Evaluating codes:** Authors 1 and 2 independently coded 1/3 (randomly-selected) of the responses for each question using the initial codebook, agreeing to identify 'mentions' rather than giving a single code to each response. Codes and descriptions were then refined by discussing disagreements.
- (3) **Coding full data set:** Authors 1 and 2 separately re-coded all responses with the updated codebook and rules.
- (4) **Defining themes:** Authors reviewed final coding and identified similarities to allow thematic grouping. We did not calculate survey inter-rater reliability because codes were not the final outcome of our analysis [42].

<sup>2</sup>The response given by this participant is an internet meme that has previously been discussed as aggressive/transphobic [30] and is not reported on further in this work.



## 4.1 Viewing Frequency

Participants reported using a variety of devices to watch scripted entertainment: Television (78 participants) and Smart TV (42), Laptop (80), Smartphone (60), Desktop PC/Computer (39), Tablet (24), iPad (29), Overhead Projector (7), Other (4), and Not Given (1). The most common device participants reported to use was either a TV or a Smart TV, collectively accounting for 32% of the reported devices.

Participants reported using a variety of services to watch scripted entertainment: Netflix (82 participants), Amazon Video (45), Virgin Media (14), Freeview (14), NowTV (13), Terrestrial TV (12), Sky (8), Apple TV (6), Sky Go (4), Not Given (2). Additionally, participants reported if they used any of the following: 'On Demand TV' (59 participants), 'Recorded Programmes' (25), 'On the Go Live TV' (7).

Participants reported varying frequencies of watching scripted entertainment, with 50% of participants reporting that they watch scripted entertainment every day: Every day (51), every other day (34), once a week (5), once every 2 weeks (3), seldom (8), never (1). There was also variety in the number of hours that they reported watching, with the majority (55%) reporting watching for 1-2 hours each day: Less than an hour (21 participants), 1-2 hours (57), 3-4 hours (22), 5 or more hours (2).

## 4.2 Subtitle Usage

Participants were asked if they regularly watch scripted entertainment with subtitles turned on: Yes (69 participants), No (32), Not Given (1). Participants reported their frequency of watching scripted entertainment with subtitles ON: Daily (36 participants), 2-3 times a week (21), Once a week (10), 1-2 times per month (18), 1-2 times per year (9), Never (6), and Not Given (2). Participants also reported frequency of watching scripted entertainment with subtitles OFF: Daily (35 participants), 2-3 times a week (24), Once a week (11), 1-2 times per month (14), 1-2 times per year (3), Never (14), and Not Given (2).

Participants reported the reasons that they watched scripted entertainment with subtitles turned on, selecting all choices that applied: Helps me understand context (37 participants), Native language translation (35), Noisy viewing conditions (34), Media content has low sound quality (33), Use subtitles to reinforce language (30), Trouble understanding international accents (26), Trouble understanding regional accents (23), Quiet viewing conditions (22), Trouble understanding national accents (16), I have a hearing loss (14), Busy using another device (10), Other (18).

**4.2.1 Subtitles to Assist in Understanding.** Participants also reported if they ever needed to pause or stop watching a show because they could not understand what was being said on screen, 55% responded that they had experienced this problem (57 participants), No (29), Maybe (15), Not Given (1).

**1) Personal Accessibility Factors:** Participants described barriers to watching content due to accessibility issues. Most commonly this was due to participants, such as P15, stating that they *"can't hear properly...and there's no subtitles so I couldn't understand them..."*. Although TV access for people with hearing loss has improved, there is still content that remains unwatchable due to a lack of subtitles, poor quality subtitles, and excessive background noise [48].

**2) Accent Challenges:** Participants reported if they had ever had trouble understanding an accent on a programme. The majority

of participants stated that they experienced this problem: Yes (63 Participants), No (37), Not Given (2). Participants described issues with accents belonging to specific people, actors, and characters. For instance, P9 described that *"Game of Thrones has some characters, which are very hard to understand"*. Furthermore, other participants mentioned specific speakers being difficult to understand such as P97 who commented that *"accents like the one Big Narstie has, grime type"* were difficult, and P37 stated they *"watched a Netflix film with Charlie Hunnam [and] didn't understand one bit of dialogue"*.

Participants also described challenges with accents specific to individual countries. A wide variety of accents were mentioned such as British by P13, or any accent different to their own such as P19 who reported *"...difficulty with international accents, regardless of country...because I don't hear them as often."*

**3) Content Barriers:** There were 28 mentions of barriers within content that led to participants needing to pause or stop watching the content. The most common barrier discussed focused on understanding actors speaking, such as P89 who reported they *"...couldn't understand what the actors were saying (in my native language) so I turned on the subtitles and rewatched that sequence."* Both P74 and P22 mention the production quality being an issue. P74 describes that they have difficulty in understanding *"Mumbling actors or bad sound! I often rewind to catch the sentence correctly..."* and P22 stating that *"sound quality or accents may be an issue."*

**4.2.2 Subtitles to Assist in Context.** Participants reported if they turned on subtitles for specific types of content, with 53% of participants reporting that they do (55 participants), 44% responded they did not (45), and two participants not responding.

**1) Context-specific Content:** Participants described content where using subtitles provided additional context. For example, P62 described using subtitles for *"quiz shows when the questions are asked very quickly"*, and P97 described using subtitles for *"Educational and difficult subjects where [they] might encounter new words/expressions"*.

**2) Foreign Language Content:** There were 25 mentions of participants using subtitles when watching content in a foreign language. For example, P102 described that they *"prefer to watch foreign shows in their native audio language, accompanied by English subtitles"*.

**3) Viewing Environment:** There were 30 mentions of external factors leading to participants needing to pause or stop watching content because they missed dialogue. P87 describes how they *"multitask often and [they] have poor attention span."* P98 stated that *"Volume too variable...didn't want to make it louder (kids sleeping)"*. P42 discussed similar problems, such as when watching content on their commute and *"...sometimes the train gets too loud and [they] need to rewind and put on subtitles."*

## 4.3 Summary of Questionnaire Findings

Our survey findings demonstrate challenges that result in viewers actively enabling subtitles. Participants described interrupting their viewing to enable subtitles due to specific accents, actors, or content types. This is echoed by press articles that criticise actors mumbling [13, 35], with this coined *mumblegate* in the UK [27]. Whilst having subtitles on all of the time could resolve these issues, our participants reported that they did not use this approach. This may be due to the impact that subtitles have on viewing experience, summarised in our Related Work. Our survey participants were heavy

watchers of content whereas viewers outside of this demographic may have different reasons for using subtitles. Furthermore, our survey was conducted in early 2019 prior to the Covid-19 Pandemic. As such, the frequency data we report around viewing and subtitle usage may not be representative of the current population.

We now explore an approach that can be used to transform subtitles from a binary (i.e., on or off) interaction to a personalised and adaptable experience that takes into account our survey results.

## 5 ADAPTIVE SUBTITLES

Our questionnaire findings highlight that people interact with subtitles based on environmental factors, challenges in content, and personal factors (e.g., tiredness, and hearing loss). A key area that was discussed by participants surrounded interacting with subtitles due to challenges in understanding specific accents and actors, and alterations for specific content types. To investigate the potential that offering real-time personalisation of subtitles would have on these areas we introduce a new subtitling approach.

Adaptive Subtitles is a media personalisation approach that can alter subtitle presentation on a subtitle-block level, contrasting with the content level approach that is currently used. Our system is based on the principle of Object Based Media (OBM) [3] where content is retained as component parts, rather than a rendered finished artefact, and delivered separately to the viewer. This increases opportunities for adaptation and personalisation based on user needs and the context of use. OBM has previously been used to allow viewers to explore music at live events [52], recap on missed television episodes [20], and to enhance audio mixes [6].

Instead of subtitles being viewed as a single object within media, we propose that additional metadata is added to subtitles files, which is then used to create opportunities for adaptation surrounding words (e.g., names, locations), phrases (e.g., catchphrases), speakers/characters, accents, audio-descriptive elements, and scene composition. This moves subtitles from a single object to a structured set of atomic elements following the guiding principles of OBM.

### 5.1 Implementation

Our survey findings demonstrate significant issues in understanding speakers due to content or accent challenges. We use this to motivate our implementation of an adaptive subtitle system that provides the viewer control over which characters have subtitles enabled/disabled. We chose this to take advantage of the working-memory that is used when consuming subtitled content and the viewer correlation that must take place between subtitle track and on-screen speaker [37]. To test our concept we created a system that consists of: 1) A second screen Adaptive Subtitles Controller Interface for controlling speaker subtitles, 2) Adaptive Subtitles TV Interface for viewing media content and rendering speaker subtitles, and 3) A nodeJS server instance to support communication between interfaces. The system architecture used [ExpressJS](#), [Angular](#), and [NodeJS](#), with [socket.io](#) for real time communication. Sample code is included in supplementary material.

**1) Adaptive Subtitles Controller Interface:** Our Adaptive Subtitles controller interface allowed control over whether subtitles were on/off for individual speakers within content. The interface consisted of play and pause buttons for the content, and speaker

cards with an image and corresponding speaker/character name that gave control over individual speakers (as shown in Figure 1.A). Cards were colour coded to match the subtitle text colour of each speaker (as shown in Figure 1.A and 1.B) and designed to work with portrait and landscape display options. The colours used for our subtitles is typical for terrestrial broadcast in the UK and follows guidance provided by OfCom [48] and BBC [16].

The controller interface displays a set of cards that correspond to each speaker in a clip. All speakers are initially in the ‘off’ state, and the images were given slight opacity to signify this to users [23]. To show/hide subtitles for a particular speaker, the user taps on the respective speaker’s image sending a `socket.emit()` event to the server. This triggers the TV interface to update style settings for the respective character resulting in subtitles being shown/hidden.

**2) Adaptive Subtitles TV Interface:** Our Adaptive Subtitles TV interface allowed viewers to only see subtitles for the specific speakers enabled on the Adaptive Subtitles controller. The interface consisted of a video window with an overlaid subtitle container at the bottom middle of the display (shown in Figure 1.B). Subtitles were styled to match BBC Guidelines [16] and styling preferences [15].

Subtitles typically have a transparent black background to assist with text contrast [16]. In a traditional web video player with subtitle support, subtitles are contained within an element (e.g., a `<div>`) that surrounds the entire subtitle block with the transparent black background being applied to this. In our Adaptive Subtitles implementation we styled individual speaker `<span>` tags to have the transparent black background (i.e. not the overall subtitle container) and used the CSS `visibility:hidden` style opposed to `display:none` to preserve subtitle placement. All noises included in subtitle tracks were unaltered and presented without a black background to differentiate them from speaker text.

WebVTT (Web Video Text Tracks) files were manually coded for each speaker by adding `<v.char>` tags, demonstrated in W3C Recommendations [53]. The edited WebVTT files could then be parsed by our Adaptive Subtitles application and inserted into HTML `<span>` elements to make them easily readable within the Document Object Model (DOM). When a `socket.emit()` event was sent from the controller to the Adaptive Subtitles TV interface containing a speaker’s subtitle state, the contents of this were parsed and relevant CSS style rules applied in order to enable/disable subtitles.

## 6 STUDY 2: EVALUATION OF ADAPTIVE SUBTITLES

The evaluation of Adaptive Subtitles comprised of a lab based user study where participants watched a selection of video clips while using the system, followed by a discussion of their experience.

The evaluation took place within our in-house user testing lab. We arranged the lab to resemble a living room with a sofa directly facing the television (for participant), and an armchair perpendicular to the television (for researcher). Participants used a Moto Z3 Play as the Adaptive Subtitles controller throughout the study (as shown in Figure 2). BBC report a median UK household television viewing distance of 2.63m, 5.5 times screen height (i.e. 5.5H), but also report that the median H is decreasing due to an increase in television size [46]. Participants sat 2.44m (4H) away from a 48" Samsung J5100 5 Series HD LED television. This is 19cm away from



**Figure 1:** A) Adaptive Subtitles ‘Controller’ interface displaying speakers for BBC’s ‘Would I Lie To You?’ clip; Lee Mack has been toggled to have subtitles displayed. B) Adaptive Subtitles displaying subtitles for speakers selected in A. C) Traditional subtitles showing dialogue for all speakers.

BBC reported median but within 1 SD of reported limits [46]. All sessions were video and audio recorded from three angles: immediate left and right of participant to assist with understanding responses, and behind participants to view interaction with controller.

## 6.1 Apparatus

Five short clips of content were selected from the BBC iPlayer online library ([bbc.co.uk/iplayer](http://bbc.co.uk/iplayer)). As we required the video files and raw subtitles we used the open source software ‘get\_iPlayer’ ([github.com/get-iPlayer](https://github.com/get-iPlayer)). We chose clips using the following criteria (similar to [11]): (a) content was not originally broadcast within a month of our study taking place (to reduce potential familiarity with content); (b) content did not contain offensive language, or potentially disturbing material; (c) there were extensive talking-head shots (i.e., as much spoken dialogue as possible); (d) content containing speakers with local regional accents (i.e. local to study location) was excluded; (e) excerpts were around five minutes in duration; (f) similar levels of activity and engagement across content; (g) subtitles were not superimposed on content before transcoding; (h) content contained challenges that affect people’s ability to lipread/speechread such as speakers turning away, and, facial hair [24, 25]. iPlayer subtitles are presented using EBU-TT (timed text) format [64]. We used get\_iPlayer to obtain these and subsequently convert these to SRT (SubRip) subtitles files. Aegisub was used to shift subtitle times to match clip length.

The clips used in the evaluation were:

**Would I Lie To You?**, 2017, Series 11, Episode 2, Broadcast: 27-Nov-2017, (00:01:35 – 00:06:59) – A comedy panel show. (only used for demonstration and study onboarding).

**Water Diviner**, 2014, Broadcast: 7-Jul-2019, (00:17:44 – 00:22:43) – A movie set after the Battle of Gallipoli.

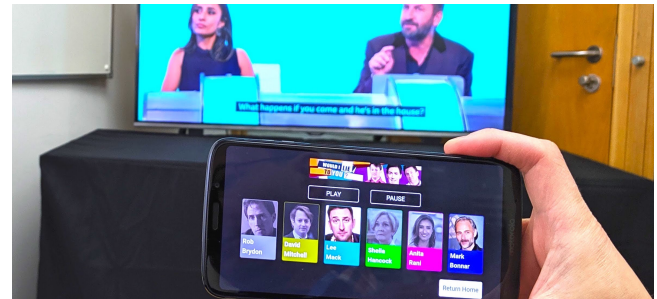
**Peaky Blinders**, 2017, Season 4, Episode 2 - "Heathens", Broadcast: 22-Nov-2017, (00:36:45 – 00:42:26) – A drama series set in England.

**A Fresh Guide to Florence with Fab 5 Freddy**, 2019, Broadcast: 27-Jul-2019, (00:29:46 – 00:35:18) – A documentary on Italian art.

**University Challenge**, 2019, Season 19/20, Episode 1, Broadcast: 15-Jul-2019, (00:02:49 – 00:05:11) – An academic quiz show.

## 6.2 Design

**Stage 1 - Questionnaire:** Participants were greeted, explained the purpose of the study, and asked to provide informed consent. A questionnaire was then given to participants, similar to the one



**Figure 2:** Evaluation setup, showing Adaptive Subtitle control interface with speaker Lee Mack selected. The Adaptive Subtitles television interface is playing the a clip and only displaying subtitles for the selected speaker.

used within our first study. The questionnaire had 14 questions across two sections. The first section contained nine questions that were used to gather demographic information; age, gender, level of education, level of computer literacy, and participants’ hearing ability. The second section contained five questions and focused on participants’ viewing experience and subtitle usage: ‘How often do you watch scripted entertainment?’, ‘How many hours per day do you watch scripted entertainment?’, ‘Do you regularly watch scripted entertainment with subtitles turned on?’, ‘How often do you watch scripted entertainment with subtitles turned ON/OFF?’.

**Stage 2 - Lab-based User Study:** Participants were asked to watch the four video clips and use our Adaptive Subtitles remote to control the subtitles on the clips to suit their own personal preferences. Participants were initially shown traditional subtitles using the ‘Would I Lie To You?’ clip. We asked participants if the volume was at a comfortable level (set at point 15 on the volume slider, ~40 Db) and if this was a typical representation of subtitles that they had used before. Participants were then shown the same clip with Adaptive Subtitles as a form of onboarding.

Participants were shown each of the four clips exclusively with the Adaptive Subtitles approach. Each clip was shown in full, and after each clip we asked questions about how participants used Adaptive Subtitles. Clip order was counterbalanced across participants using a Williams Balanced Latin Square.

**Stage 3 - Post-session Discussion:** After viewing all of the clips, we used the UX Subtitle Framework [18] to scaffold a semi-structure



interview. This framework has been used to assess the UX of different subtitle approaches [9, 10, 19], and also by media industry practitioners [5]. The framework allows overall measure of UX to be assessed when viewing different methods of subtitle display.

### 6.3 Participants

Participants were over 18 years-old and self identified as turning on subtitles whilst watching media content at least once in the last three months. We recruited 19 participants from a local university aged between 19 and 53 ( $M = 28.38$  years,  $SD = 8.77$ ). Participants were compensated with a £10 gift voucher.

We used an open text field for gender: Male = 14, Female = 5. Participants reported on their highest level of education: University (17 participants), College (1), High School (1), Other (0). Participants reported on their level of computer literacy: Excellent (15 participants), Good (4), Fair (0), and Poor (0). In total, two participants self-reported having a hearing loss. One participant had Moderate hearing loss for 26 years due to a Virus or Disease, and one had Mild hearing loss for 3 years due to exposure to loud noise. Neither reported using hearing aids or cochlear implants.

Participants reported the frequency of watching scripted entertainment: Every day (10 participants), Every other day (4), Once a week (4), Once every 2 weeks (0), Seldom (1), Never (0). There was also variety in the number of hours that they reported watching each day: <1 hour (4 participants), 1-2 hours (13), 3-4 hours (2), 5 or more hours (0). Participants were asked if they regularly watch scripted entertainment with subtitles turned on: Yes (15), No (4).

Participants reported their frequency of watching scripted entertainment with subtitles ON: Daily (6), 2-3 times a week (6), Once a week (5), 1-2 times per month (2), 1-2 times per year (0), Never (0). Participants reported frequency of watching scripted entertainment with subtitles OFF: Daily (7), 2-3 times a week (7), Once a week (1), 1-2 times per month (3), 1-2 times per year (0), Never (1).

### 6.4 Results

All sessions were transcribed and analysed by the authors. While every attempt was made to remain impartial throughout data gathering and analysis, a potential bias may exist as an author was present for all interview sessions. The use of an interview guide with structured questions reduces bias in this regard. Transcripts from all sessions were created from the experiment video files and were blocked according to the related sections within the interview guide. Sections were combined between participants and examined individually based on interview guide components. Closed descriptive coding was carried out with attention paid towards the benefits and drawbacks of traditional subtitles and Adaptive Subtitles whilst also exploring users perceptions towards real-time adaptations of media elements. Individual quotes were coded with authors agreeing on the inclusion of each within their categories. Conclusions within results are drawn from data general trends.

**6.4.1 Context-Based Adaptions.** Participants commented they would “use it [adaptive subtitles] pretty often” (P1) but that it “would depend on what I’m watching” (P7). This awareness of context-based adaptations was also highlighted by P10, stating that “if it was a movie or a TV show that was a one off, I would go for traditional [subtitles]”. In some situations verbal content is less important than on-screen

visuals and that for documentaries you “don’t need to know what everyone is saying” (P8). Participants also saw the benefit of having adaptive subtitles present within serialised content, discussing that it “would be useful to have it going across episodes” (P10) and that this would reduce the overall attention lost to initially setting the system – “you would choose a setting and then leave it” (P2).

Our participants were divided in how they approached adaptive subtitling within the University Challenge clip, with usage being very different to story-driven media. In this clip most participants turned on the subtitles for the quiz show host, “even though I can hear the questions, they are long questions so then I can read them.” (P4). This was echoed by P5 who stated that “some of the questions might be a bit technical, so it gives you some reassurance.” Turning on subtitles for only the presenter gave the added value of enabling participants to *play along* with the show itself with P4 describing that they turned subtitles off for contestants “so that it didn’t give me the answers”. Some participants took a different approach to adaptive subtitling in this content type and also turned on subtitles when teams were conferring, describing that “you would just hear whispering...but with subtitles you get it and you learn more” (P17).

**6.4.2 Within-Content Adaptions.** Participants noted that aspects such as character accents and their previous exposure to a given show impacted on why they chose to use subtitles for given characters. Participants discussed that content-based difficulties, such as understanding accents, caused them to enable adaptive subtitles for individual speakers - “I struggled with some of the accents” (P8). P18 elaborated on this by discussing that “the issue isn’t how loud they speak, it’s really the accents”. Participants found some speakers to be more challenging to understand than others, with one participant turning on adaptive subtitles for “the men...because they had stronger accents, with trying to be suspicious and all” (P14). Some participants were quick to adapt to accents that they understood with P16 stating that “it was the initial anticipation but then I realised that I could hear him fine so turned him off”. One participant commented that previous exposure to one of the shows within the study assisted in determining which characters to have subtitles on for, “I watch Peaky Blinders, I’ve listened to the accents before...I was almost pre-empted to turn them on.” (P5). This was contrasted with P12 who stated that “I’m not very familiar with the programme, which made it hard to find out which ones to put on”.

**6.4.3 Benefit I - Adaptive Subtitles Increases Focus on Main Content.** During our evaluation, participants commented that that one of the main challenges present with traditional subtitles is that “you are always drawn to the words and you might miss something” (P16). This adds to the cognitive load associated with watching content as “first you read it [subtitles], then you reinforce it with talking, and then you get it”. (P14). Using the adaptive subtitles approach, participants commented that they are “able to focus on the clip itself, there is less to read...I’m only reading what I want to read” (P5). Participants also discussed that adaptive subtitles “helps you focus on what is needed.” (P2), and that this approach doesn’t “distract me from what is going on as much as traditional subtitles”.

**6.4.4 Benefit II - Adaptive Subtitle Presentation alters Content Consumption Method.** A common view from participants focused on the disconnect that traditional subtitles create with P18



stating that *“if you have subtitles on you are slightly disconnected with what is going on, I’m always concentrating on the subtitles”*. This was echoed by P14 who reflected on moving between subtitled and non-subtitled content, describing *“I was able to be there and see the pictures and felt like I was zooming in and being there while when the other guy was talking and the subtitles come on I feel like I’m shut out. Its a barrier between me and what is happening”*.

A consequence of our subtitle styling approach (see Implementation), is that in some situations a visual ‘gap’ between subtitle blocks appears on the screen. Whilst this was not something that we had intended to be an aesthetic feature of Adaptive Subtitles, it is something that participants acknowledged within discussion. Participants commented that they *“knew that I shouldn’t be reading it all at once”* (P14) and when the gap was present they *“didn’t read the second statement, I was able to wait”* (P13). Despite one participant finding this feature to be *“offputting”* (P7), participants commented that this method of subtitle display altered the method in which they consumed subtitled content and that *“you almost leave space in your mind waiting for the reply”* (P5).

**6.4.5 Trade-Off I - Personalising Content Leads to Increased Physical and Mental Effort.** Whilst participants were comfortable with the concept of adaptive subtitles, they saw clear disadvantages in the effort required to create a customised list of characters that would have subtitles enabled. P13 described this as *“a very involved process”* and P8 added to this by commenting that *“I felt a bit less involved because I was doing a task and doing something instead of just sitting back and watching”*. P13 raised concerns about its usage in some shows, stating that *“...for dramas, its moving between the fictional world and the real world. Something like pressing pause is a conscious choice and when you press play you are going back in, with this it is like you’re never really getting in”*. Participants commented on the overall usability of Adaptive Subtitles, saying that it *“felt like I had to do more work, it was more effort on my part”* (P4). P2 added to this by discussing that the implications of this challenge scales with the number of characters present, and that *“if there are many characters, matching the object to the person on the screen is hard”*. Despite this challenge, P9 highlighted the trade-off that has to happen when personalising media content, stating that our approach was *“easy to use, but there is more to use”*.

**6.4.6 Trade-Off II - Second Screen Device Interaction Alters the Passive Media Experience.** Our adaptive subtitles implementation was facilitated through a second screen application that allowed participants to individually choose which characters had subtitles enabled. Participants felt that this approach altered the overall experience, *“TV is a very passive thing, you want gentle actions...the remote is more involved it turns it into an active experience”* (P13). Participants commented on the trade offs of this approach, with P7 suggesting that they felt *“more involved in terms of what was happening in what was being said, but less involved because I had to look at the remote”*. Challenges in using this second screen device was also discussed by other participants, with P10 noting that *“one problem is that you have to look down and look up, you might be missing content that is on the screen, [but] its only a short lapse in concentration”*. In our developed application we matched up the background colour of characters on the second screen device with their individual subtitle colours on the main display. This was

carried out in order to assist with the move from selection of adaptive subtitles to the consumption of adaptive subtitled content and the move between devices that is part of this. The use of multiple colours in subtitles is a common feature across UK terrestrial TV [16, 50] but less common in other countries and in online platforms. Participants commented that our approach meant that *“you already know what colour what character is”* (P18) and that they *“appreciated the colour coding so I could tell who was talking”* (P14). The consistent application of colour across devices (see Figure 1) *“help(s) with contextual understanding of characters and their names. When the colours started I was able to marry up who these people were”* (P3).

## 6.5 Study Limitations

Our evaluation of Adaptive Subtitles focused on short, lab based exposure. Participants viewed four clips in a short amount of time, and as such, we are only able to generalise our findings to this “setup” period. The nature of participant exposure to Adaptive Subtitles in our work focused on initial system usage, and we did not experience the set and forget phenomenon [19] that would be expected over longer usage. We encountered issues surrounding split-attention due to the use of our second screen remote, echoing challenges discussed in other work [8]. However, the on-boarding of users to new technology and concepts should be carefully planned. Our work is a necessary step in understanding challenges in this area. Our approach requires additional effort to edit content into atomic elements, like other OBM systems [6, 20, 52]. However, in some cases this is done automatically when colour is added to subtitles by broadcasters, therefore it only requires effort to match each colour to the first instance of the speaker. Our current approach would only work on a digital web-based display system (e.g., Netflix).

## 7 DISCUSSION

Giving users the ability to personalise the way they experience content is challenging. Any time that is spent implementing an adaption is time that is not spent consuming the media itself. The task that users’ go through to adapt an interface follows the same, broad, iterative process: *consuming content* → *deciding that content should be adapted* → *selecting content to be adapted* → *evaluating if the adaption is acceptable*. The added challenge with real-time media adaptations is that there is a greater emphasis on ensuring that the final three aspects in this cycle take as limited amount of time as possible. We reflect on this and present design considerations to enable others to better understand the interaction challenges that were faced in this work. We initially discuss these as collective guidance and subsequently expand on these individually.

Providing granular control over content allows viewers to move between understanding content and being immersed in scene context, but this complexity can overwhelm users. To assist with this complexity, there should be an onboarding period for the content being watched, with this viewed as separate to onboarding of the technology itself. This onboarding period can assist in reducing information overload, but can initially be seen as a distraction during viewing. Despite this, the short-term distraction created should lead to long-term benefits for viewers in terms of improved viewing experience. In our study, users had to do carry out this process at the start of short clips. Most use-cases for our approach would

be when watching longer content: 45 min TV episode, a 120 min movie, or serialised content that could span several episodes.

### 7.1 Understanding vs. Immersion

In our work, participants commented that they removed subtitles within shows where the visuals are more important than the information being described to assist in promoting involvement in the show. In these situations subtitles were important for understanding content but are seen as distracting during establishing shots when voice-over was being used, leading to a reduction in how viewers experience the context of a particular shot. The removal of subtitles during these instances assisted in making users feel more immersed. Conversely, participants described that they added subtitles within shows where they struggled to understand particular characters. In these cases, the lack of subtitles would lead to little understanding of the content and subsequently a reduction in overall contextual experience.

Participants had very different reasons for their individual setup options for adaptive subtitles. Each attempted to find their own sweet spot for understanding content whilst also creating adaptations to allow immersion in the viewing experience. In our work, we found that participants added subtitles to content when they wanted to understand content but removed subtitles when they wanted to feel more immersed in the show. The *understanding content* → *immersive context* continuum that participants were interacting with is a careful balancing act that can change often, even between scenes. Giving viewers the ability to move along this highlights the challenge in creating adaptive interfaces for real-time media.

Instead of assuming what type of adaption will be required, we recommend designers should **provide users with the ability to create granular adaptations to move between understanding content and being immersed in a specific context**. This provides viewers with the opportunity to decide how they wish to consume content and may lead to positive alterations in lean-back or lean-forward experiences [54] based on their own viewing goals.

### 7.2 Technology and Content Onboarding

In our evaluation, participants took part in two distinct onboarding experiences. Firstly, they were introduced to the concept of adaptive subtitles and our implementation of this method (i.e. onboarding to the technology). Secondly, they decided on the adaptations they would like to make to a given clip before consuming the media (i.e. onboarding to the content). These adaptations to clips took place when participants viewed them as being necessary and transitioned from onboarding to real-time adaptations.

The personalisation of media content requires focus and attention from viewers. This shift in attention from the media itself to the media controls was challenging for our participants. They discussed how it was a complex process that involved identifying the character on the screen, locating that character on a secondary device, and finally selecting subtitle state. We describe during the introduction to our discussion the broad, iterative, process that users go through when making adaptations. On-boarding alters this process in that it produces an opportunity to decide, select, and evaluate adaptations in a situation where the consumption of media is no longer the primary objective.

One method that television shows use to create onboarding experiences is by using episode recaps that provide viewers with information regarding ongoing plot lines and important characters [41]. Similar techniques could be used to on-board viewers when interacting with real-time personalisation of media content. This would allow for adaptations to content in a situation where consumption of the media is less important, or carried out based on previously altered content (e.g. between episodes in serialised content or for common actors). This creates a clear separation between onboarding to adaptations for particular content types and the onboarding of how the technology works. We recommend that designers should **consider the onboarding of viewers to content and the technology as separate elements and cater for these using different techniques**.

### 7.3 Trade-Offs and Benefits

During the evaluation, participants found it challenging to match up characters between our second screen interface and the content on the main screen. This could be due to the length of our study clips (5 minutes) compared to entire television show episodes (30–60 minutes) and feature films (120+ minutes). Despite the short clip lengths, participants commented that they saw the potential benefits of using this when watching longer content.

The approaches that our participants used to determine which speakers to enable were related to the coping strategies discussed by participants in our initial survey. For instance, some participants turned on speakers they identified had heavy or unfamiliar accents. Participants commented that once their subtitle view (and therefore overall content) was personalised they found it to be less distracting than their previous experiences using traditional subtitles. This reduced the level of disconnect between themselves and the media. Participants acknowledged the trade-off between short-term distraction at the start of a piece of content, versus long term benefits of a personalised viewing experience. This process follows the set-and-forget phenomenon [19]. We recommend that designers should **embrace the set-and-forget phenomenon when developing real-time media adaptations** to improve viewer involvement with content over the long-term. This allows viewers to focus on personalising their experience during points where story elements are limited rather than during key points of content and will lead to increased levels of engagement.

## 8 CONCLUSION

Subtitles are commonly thought to be an accessibility feature, and are traditionally viewed to be only used by people with a hearing impairment. However, for 80% of subtitle users this is not the case [49]. To understand how and why viewers adapt their subtitle usage, we conducted an online questionnaire with 102 subtitle users. Our participants reported using subtitles based on the language being spoken, accents, scene-context, and programme quality.

Inspired by the challenges that our participants faced, and recent developments in media production [20], we introduced a new subtitling approach called Adaptive Subtitles to investigate the challenges and opportunities that exist when interacting with real-time personalisation of subtitled media content. Our evaluation illustrated that the personalisation needs of an individual changes

based on what they are watching and how they wish to consume it. For example, people may turn on subtitles for individual characters in a movie due to challenges in understanding accents, and in quiz shows subtitles can be turned off for contestants so people can more easily play along. We also consider content of different lengths. For example, the benefits of using adaptive subtitles on long form content (e.g., a movie) likely outweigh the drawbacks. Whereas the drawbacks of using it on a short term content (e.g., a short TV episode) may outweigh the benefits.

We propose three design considerations that should be used when developing media personalisation features: 1) Provide users with the ability to create granular adaptations to move between understanding content and experiencing scenes in context, 2) Consider the onboarding of viewers to content and the technology as separate elements and cater for these using different techniques, and 3) Embrace the set-and-forget phenomenon when developing real-time media adaptations. We suggest that by following these recommendations should increase engagement with media content.

## REFERENCES

- [1] M. Armstrong. 2017. Automatic Recovery and Verification of Subtitles for Large Collections of Video Clips. *SMPTE Motion Imaging Journal* 126, 8 (2017), 1–7. <https://doi.org/10.5594/JMI.2017.2732858>
- [2] M. Armstrong, S. Bowman, M. Brooks, A. Brown, J. Carter, A. Jones, M. Leonard, and T. Preece. 2020. Taking Object-Based Media from the Research Environment Into Mainstream Production. *SMPTE Motion Imaging Journal* 129, 5 (2020), 30–38. <https://doi.org/10.5594/JMI.2020.2990255>
- [3] Mike Armstrong, Matthew Brooks, Anthony Churnside, Michael Evans, Frank Melchior, and Matthew Shotton. 2014. Object-based broadcasting-curation, responsiveness and user experience. *IBC2014 Conference* 2014, 1 (2014), 1–8. <https://doi.org/10.1049/ib.2014.0038>
- [4] Mike Armstrong, Andy Brown, Michael Crabb, Chris J. Hughes, Rhianne Jones, and James Sandford. 2016. Understanding the Diverse Needs of Subtitle Users in a Rapidly Evolving Media Landscape. *SMPTE Motion Imaging Journal* 125, 9 (2016), 33–41. <https://doi.org/10.5594/JMI.2016.2614919>
- [5] BBC Research and Development. 2016. Subtitle Quality - Measuring and improving subtitle quality. [www.bbc.co.uk/rd/projects/live-subtitle-quality](http://www.bbc.co.uk/rd/projects/live-subtitle-quality)
- [6] BBC Research and Development. 2019. Casualty, Loud and Clear. <https://www.bbc.co.uk/rd/blog/2019-08-casualty-tv-drama-audio-mix-speech-hearing>
- [7] Tim Brooks. 2019. Television and Record Industry Nielson Ratings. <https://timbrooks.net/ratings/>
- [8] Andy Brown, Amaia Aizpurua, Caroline Jay, Michael Evans, Maxine Glancy, and Simon Harper. 2019. Contrasting delivery modes for second screen TV content. Push or Pull? *International Journal of Human-Computer Studies* 129 (2019), 15–26. <https://doi.org/10.1016/j.ijhcs.2019.03.007>
- [9] Andy Brown, Rhia Jones, Mike Crabb, James Sandford, Matthew Brooks, Mike Armstrong, and Caroline Jay. 2015. Dynamic Subtitles: The User Experience. In *Proceedings of the ACM International Conference on Interactive Experiences for TV and Online Video* (Brussels, Belgium) (TVX '15). ACM, New York, NY, USA, 103–112. <https://doi.org/10.1145/2745197.2745204>
- [10] Andy Brown, Jayson Turner, Jake Patterson, Anastasia Schmitz, Mike Armstrong, and Maxine Glancy. 2017. Subtitles in 360-degree Video. In *Adjunct Publication of the 2017 ACM International Conference on Interactive Experiences for TV and Online Video* (Hilversum, The Netherlands) (TVX '17 Adjunct). ACM, New York, NY, USA, 3–8. <https://doi.org/10.1145/3084289.3089915>
- [11] Denis Burnham, Greg Leigh, William Noble, Caroline Jones, Michael Tyler, Leonid Grebennikov, and Alex Varley. 2008. Parameters in television captioning for deaf and hard-of-hearing adults: Effects of caption rate versus text reduction on comprehension. *Journal of deaf studies and deaf education* 13, 3 (2008), 391–404. <https://doi.org/10.1093/deafed/enn003>
- [12] Cristina Cambra, Olivier Penacchio, N ria Silvestre, and Aurora Leal. 2014. Visual attention to subtitles when viewing a cartoon by deaf and hearing children: an eye-tracking pilot study. *Perspectives* 22, 4 (2014), 607–617. <https://doi.org/10.1080/0907676X.2014.923477>
- [13] Jessica Carpani. 2019. BBC criticised for ‘mumbling’ adaptation of A Christmas Carol. <https://www.telegraph.co.uk/news/2019/12/23/bbc-criticised-mumbling-adaptation-christmas-carol/>
- [14] Brad Chisholm. 1987. Reading Intertitles. *Journal of Popular Film and Television* 15, 3 (1987), 137–142.
- [15] British Broadcasting Corporation. 2018. BBC Global Experience Language. <https://www.bbc.co.uk/gel>
- [16] British Broadcasting Corporation. 2018. BBC Subtitle Guidelines. <http://bbc.github.io/subtitle-guidelines/>
- [17] Michael Crabb, Michael Heron, Rhianne Jones, Mike Armstrong, Hayley Reid, and Amy Wilson. 2019. Developing Accessible Services: Understanding Current Knowledge and Areas for Future Support. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland UK) (CHI '19). ACM, New York, NY, USA, Article 216, 12 pages. <https://doi.org/10.1145/3290605.3300446>
- [18] Michael Crabb, Rhianne Jones, and Mike Armstrong. 2015. The Development of a Framework for Understanding the UX of Subtitles. In *Proceedings of the 17th International ACM SIGACCESS Conference on Computers & #38; Accessibility* (Lisbon, Portugal) (ASSETS '15). ACM, New York, NY, USA, 347–348. <https://doi.org/10.1145/2700648.2811372>
- [19] Michael Crabb, Rhianne Jones, Mike Armstrong, and Chris J. Hughes. 2015. Online News Videos: The UX of Subtitle Position. In *Proceedings of the 17th International ACM SIGACCESS Conference on Computers and Accessibility* (Lisbon, Portugal) (ASSETS '15). Association for Computing Machinery, New York, NY, USA, 215–222. <https://doi.org/10.1145/2700648.2809866>
- [20] Michael Evans, Tristan Ferne, Zillah Watson, Frank Melchior, Matthew Brooks, Phil Stenton, and Ian Forrester. 2016. Creating object-based experiences in the real world. *Proceedings of the IBC Conference 2016* 2014, 1 (2016), 1–8. <https://doi.org/10.1049/ibc.2016.0034>
- [21] Wendy Fox. 2016. Integrated titles: An improved viewing experience? *Eye tracking and Applied Linguistics* 2 (2016), 5. <https://doi.org/10.17169/langsci.b108.233>
- [22] Alejandra Garrido, Sergio Firmenich, Gustavo Rossi, Julian Grigera, Nuria Medina-Medina, and Ivana Harari. 2012. Personalized web accessibility using client-side refactoring. *IEEE Internet Computing* 17, 4 (2012), 58–66. <https://doi.org/10.1109/MIC.2012.143>
- [23] Google. 2019. Material Guidelines - Displaying State. <https://material.io/design/interaction/states.html>
- [24] Benjamin M. Gorman and David R. Flatla. 2017. A Framework for Speechreading Acquisition Tools. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) (CHI '17). Association for Computing Machinery, New York, NY, USA, 519–530. <https://doi.org/10.1145/3025453.3025560>
- [25] Benjamin M. Gorman and David R. Flatla. 2018. MirrorMirror: A Mobile Application to Improve Speechreading Acquisition. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) (CHI '18). Association for Computing Machinery, New York, NY, USA, 1–12. <https://doi.org/10.1145/3173574.3173600>
- [26] Vicki L. Hanson and John T. Richards. 2003. A Web Accessibility Service: Update and Findings. *SIGACCESS Access. Comput.* 2003, 77–78 (Sept. 2003), 169–176. <https://doi.org/10.1145/1029014.1028661>
- [27] Ellie Harrison. 2017. Why is it so hard to hear the dialogue in TV dramas? <https://www.radiotimes.com/news/tv/2017-02-23/tv-sound-problems-drama-dialogue/>
- [28] Chris Hughes, Mario Montagud Climent, and Peter tho Pesch. 2019. Disruptive Approaches for Subtitling in Immersive Environments. In *Proceedings of the 2019 ACM International Conference on Interactive Experiences for TV and Online Video* (Salford (Manchester), United Kingdom) (TVX '19). Association for Computing Machinery, New York, NY, USA, 216–229. <https://doi.org/10.1145/3317697.3325123>
- [29] Chris J. Hughes, Mike Armstrong, Rhianne Jones, and Michael Crabb. 2015. Responsive Design for Personalised Subtitles. In *Proceedings of the 12th Web for All Conference* (Florence, Italy) (W4A '15). ACM, New York, NY, USA, Article 8, 4 pages. <https://doi.org/10.1145/2745555.2746650>
- [30] Samantha Jaroszewski, Danielle Lottridge, Oliver L. Haimson, and Katie Quehl. 2018. "Genderfluid" or "Attack Helicopter": Responsible HCI Research Practice with Non-Binary Gender Variation in Online Communities. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) (CHI '18). Association for Computing Machinery, New York, NY, USA, 1–15. <https://doi.org/10.1145/3173574.3173881>
- [31] Carl J. Jensema, Ramalinga Sarma Danturthi, and Robert Burch. 2000. Time spent viewing captions on television programs. *American annals of the deaf* 2000, 1 (2000), 464–468. <https://doi.org/10.1353/aad.2012.0144>
- [32] Helen Katz. 2006. *The media handbook: A complete guide to advertising media selection, planning, research, and buying*. Routledge, Oxford, UK. <https://doi.org/10.4324/9781315537870>
- [33] Jan-Louis Kruger, Stephen Doherty, and Maria-T Soto-Sanfiel. 2017. Original Language Subtitles: Their Effects on the Native and Foreign Viewer. *Communicar: Media Education Research Journal* 25, 50 (2017), 23–32. <https://doi.org/10.3916/C50-2017-02>
- [34] Jan-Louis Kruger, Agnieszka Szarkowska, and Izabela Krejtz. 2015. Subtitles on the moving image: an overview of eye tracking studies. *Refractory: A Journal of Entertainment Media* 25 (2015), 1–14. <http://hdl.handle.net/1959.14/1040614>
- [35] Michael Lallo. 2017. Speak up! How ‘mumble acting’ is ruining TV and film. <https://www.smh.com.au/entertainment/tv-and-radio/speak-up-how-mumble-acting-is-ruining-tv-and-film-20170201-gu2u5j.html>



- [36] M Lambooi, MJ Murdoch, Wijnand A IJsselstein, and Ingrid Heynderickx. 2013. The impact of video characteristics and subtitles on visual comfort of 3D TV. *Displays* 34, 1 (2013), 8–16. <https://doi.org/10.1016/j.displa.2012.09.002>
- [37] Mina Lee, Beverly Roskos, and David R. Ewoldsen. 2013. The Impact of Subtitles on Comprehension of Narrative Film. *Media Psychology* 16, 4 (2013), 412–440. <https://doi.org/10.1080/15213269.2013.826119> arXiv:<https://doi.org/10.1080/15213269.2013.826119>
- [38] Margaret S Jelinek Lewis. 2000. Television captioning: A vehicle for accessibility and literacy. *On-line Proceedings of CSUN 2000*, 1 (2000), 1–5.
- [39] Lluís Manchón and Pilar Orero. 2018. Usability tests for personalised subtitles. *Translation Spaces* 7, 2 (2018), 263–284. <https://doi.org/10.1075/ts.18016.man>
- [40] Paul Markham. 1999. Captioned videotapes and second-language listening word recognition. *Foreign Language Annals* 32, 3 (1999), 321–328. <https://doi.org/10.1111/j.1944-9720.1999.tb01344.x>
- [41] Shaun Patrick McCarthy, Yaron Sole, Trevor James Walker, Arun Velayudhan Pillai, and Venkatraman Prabhu. 2020. Personalized recap clips. US Patent 10,555,023.
- [42] Nora McDonald, Sarita Schoenebeck, and Andrea Forte. 2019. Reliability and inter-rater reliability in qualitative research: Norms and guidelines for CSCW and HCI practice. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 1–23.
- [43] Holger Mitterer and James M. McQueen. 2009. Foreign Subtitles Help but Native-Language Subtitles Harm Foreign Speech Perception. *PLOS ONE* 4, 11 (11 2009), 1–5. <https://doi.org/10.1371/journal.pone.0007785>
- [44] Kyle Montague, Vicki L. Hanson, and Andy Cogley. 2012. Designing for Individuals: Usable Touch-screen Interaction Through Shared User Models. In *Proceedings of the 14th International ACM SIGACCESS Conference on Computers and Accessibility* (Boulder, Colorado, USA) (ASSETS '12). ACM, New York, NY, USA, 151–158. <https://doi.org/10.1145/2384916.2384943>
- [45] Netflix. 2019. Netflix Subtitle Preferences. [netflix.com/subtitlepreferences](https://netflix.com/subtitlepreferences)
- [46] K Noland and L Truong. 2015. A survey of UK television viewing conditions. *BBC Research & Development White Paper* 287 (2015), 1–58. <https://www.bbc.co.uk/rd/publications/whitepaper287>
- [47] Don Norman, Jim Miller, and Austin Henderson. 1995. What You See, Some of What's in the Future, and How We Go About Doing It: HI at Apple Computer. In *Conference Companion on Human Factors in Computing Systems* (Denver, Colorado, USA) (CHI '95). ACM, New York, NY, USA, 155–. <https://doi.org/10.1145/223355.223477>
- [48] The Office of Communications UK (Ofcom). 2018. Making on-demand services accessible. [https://www.ofcom.org.uk/\\_data/assets/pdf\\_file/0014/131063/Statement-Making-on-demand-services-accessible.pdf](https://www.ofcom.org.uk/_data/assets/pdf_file/0014/131063/Statement-Making-on-demand-services-accessible.pdf).
- [49] OfCom. 2006. Television access services review. <https://www.ofcom.org.uk/consultations-and-statements/category-1/accessservs>
- [50] OfCom. 2017. Ofcom's Code on Television Access Services. [https://www.ofcom.org.uk/\\_data/assets/pdf\\_file/0020/97040/Access-service-code-Jan-2017.pdf](https://www.ofcom.org.uk/_data/assets/pdf_file/0020/97040/Access-service-code-Jan-2017.pdf)
- [51] OfCom. 2019. Adults' media use and attitudes Report. <https://www.ofcom.org.uk/research-and-data/media-literacy-research/adults/adults-media-use-and-attitudes>
- [52] Matthew Paradis, Rebecca Gregory-Clarke, and Frank Melchior. 2015. Venue-Explorer, Object-Based Interactive Audio for Live Events. *Proceedings of the International Web Audio Conference* 2015, 1 (January 2015), 1–5.
- [53] Silvia Pfeiffer. 2019. WebVTT: The Web Video Text Tracks Format. <https://www.w3.org/TR/2019/CR-webvtt1-20190404/>.
- [54] Krishnan Ramanathan, Yogesh Sankarabramaniam, and Vidhya Govindaraju. 2011. Personalized Video: Leanback Online Video Consumption. In *Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval* (Beijing, China) (SIGIR '11). Association for Computing Machinery, New York, NY, USA, 1277–1278. <https://doi.org/10.1145/2009916.2010158>
- [55] RNID. 2016. What Happens In An Audiology Appointment. <https://rnid.org.uk/information-and-support/hearing-loss/getting-your-hearing-tested/what-happens-in-an-audiology-appointment/>.
- [56] S4C. 2001. *Research into the demand for Welsh language subtitling in Wales*. RNID, RNID Cymru. [https://www.s4c.cymru/abouts4c/corporate/pdf/e\\_adroddiad\\_isdeitlo.pdf](https://www.s4c.cymru/abouts4c/corporate/pdf/e_adroddiad_isdeitlo.pdf)
- [57] Jason M. Silveira and Frank M. Diaz. 2014. The effect of subtitles on listeners' perceptions of expressivity. *Psychology of Music* 42, 2 (2014), 233–250. <https://doi.org/10.1177/0305735612463951> arXiv:<https://doi.org/10.1177/0305735612463951>
- [58] David Sloan, Matthew Tylee Atkinson, Colin Machin, and Yunqiu Li. 2010. The Potential of Adaptive Interfaces As an Accessibility Aid for Older Web Users. In *Proceedings of the 2010 International Cross Disciplinary Conference on Web Accessibility (W4A)* (Raleigh, North Carolina) (W4A '10). ACM, New York, NY, USA, Article 35, 10 pages. <https://doi.org/10.1145/1805986.1806033>
- [59] David Sloan, Peter Gregor, Murray Rowan, and Paul Booth. 2000. Accessible Accessibility. In *Proceedings on the 2000 Conference on Universal Usability* (Arlington, Virginia, USA) (CUU '00). ACM, New York, NY, USA, 96–101. <https://doi.org/10.1145/355460.355480>
- [60] Peter Thompson. 2000. Notes on Subtitles and Superimpositions. *Chicago Media Works* 1, 18 (2000), 1–4.
- [61] Garreth W. Tigwell, David R. Flatla, and Rachel Menzies. 2018. It's Not Just the Light: Understanding the Factors Causing Situational Visual Impairments during Mobile Interaction. In *Proceedings of the 10th Nordic Conference on Human-Computer Interaction* (Oslo, Norway) (NordCHI '18). ACM, New York, NY, USA, 338–351. <https://doi.org/10.1145/3240167.3240207>
- [62] Garreth W. Tigwell, Benjamin M. Gorman, and Rachel Menzies. 2020. Emoji Accessibility for Visually Impaired People. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) (CHI '20). Association for Computing Machinery, New York, NY, USA, 1–14. <https://doi.org/10.1145/3313831.3376267>
- [63] Sarah J Tracy. 2019. *Qualitative research methods: Collecting evidence, crafting analysis, communicating impact*. John Wiley & Sons, Oxford, UK.
- [64] European Broadcasting Union. 2018. EBU-TT-D Subtitling Distribution Format. <https://tech.ebu.ch/docs/tech/tech3380.pdf>
- [65] T. Vigier, Y. Baveye, J. Rousseau, and P. Le Callet. 2016. Visual attention as a dimension of QoE: Subtitles in UHD videos. *2016 Eighth International Conference on Quality of Multimedia Experience (QoMEX)* 2016, 1 (June 2016), 1–6. <https://doi.org/10.1109/QoMEX.2016.7498924>
- [66] W3C. 2019. Making Audio and Video Media Accessible - Captions/Subtitles. <https://www.w3.org/WAI/media/av/captions/>
- [67] Jacob O. Wobbrock, Krzysztof Z. Gajos, Shaun K. Kane, and Gregg C. Vanderheiden. 2018. Ability-Based Design. *Commun. ACM* 61, 6 (May 2018), 62–71. <https://doi.org/10.1145/3148051>